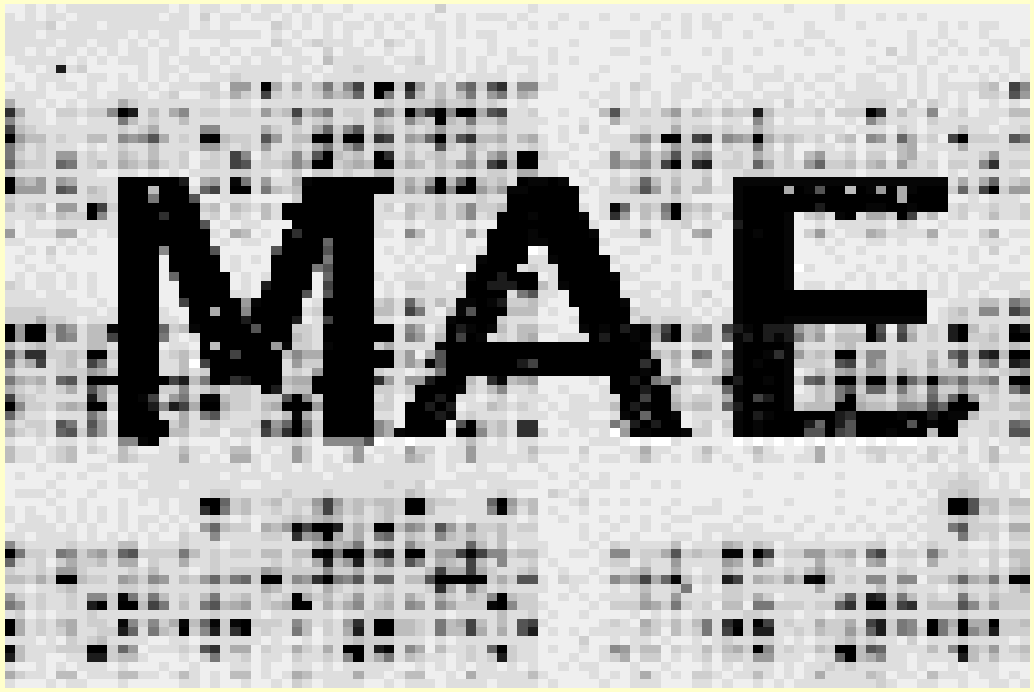
II. Examples of MicroArray Explorer

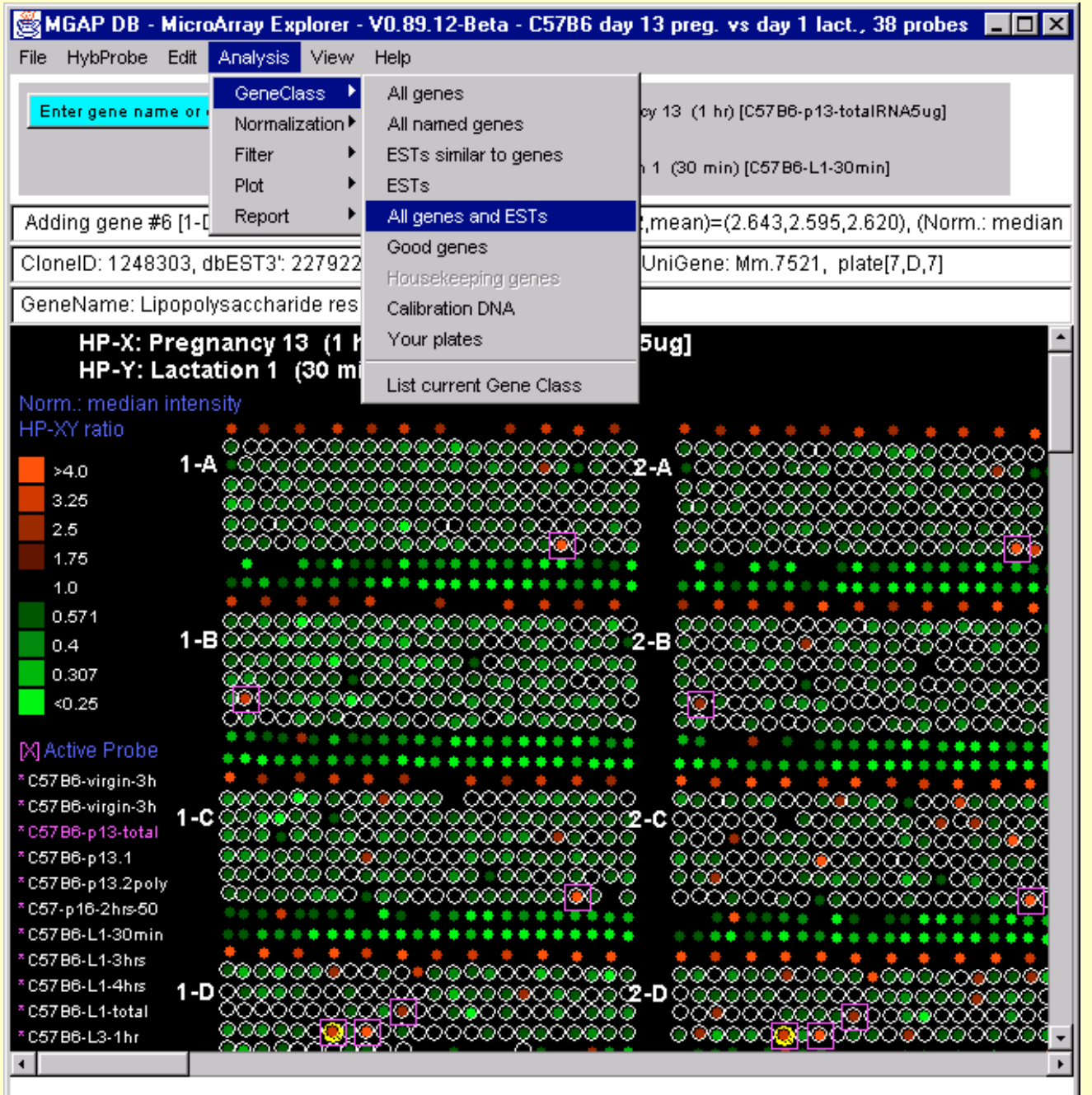


Outline

1. Description
2. Importing data
3. Examples of analysis capabilities

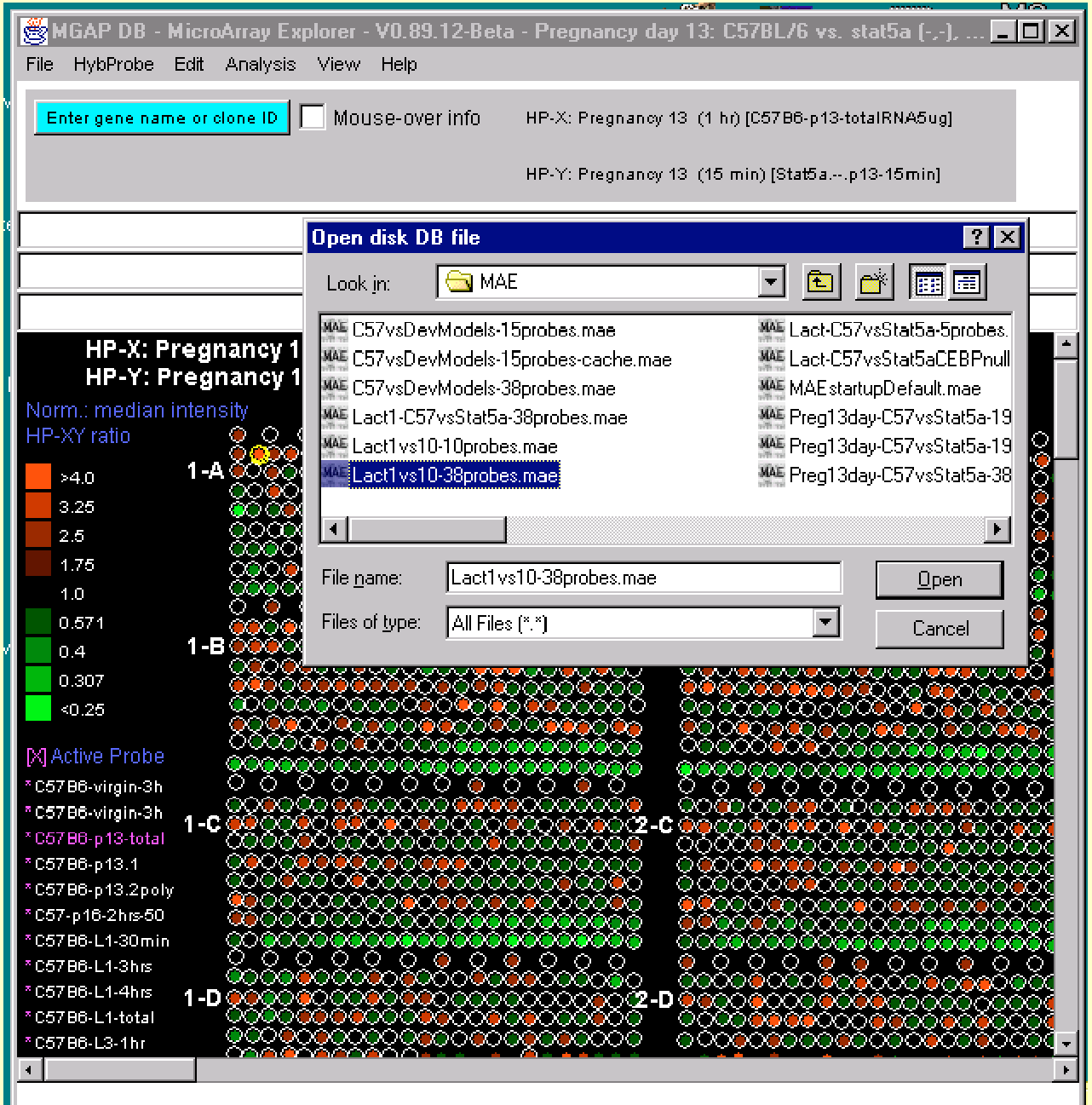
II.1 MAExplorer uses GUI-based User Interface

- Specify sets of genes for *all named genes* and *all ESTs* indicated by white circles.



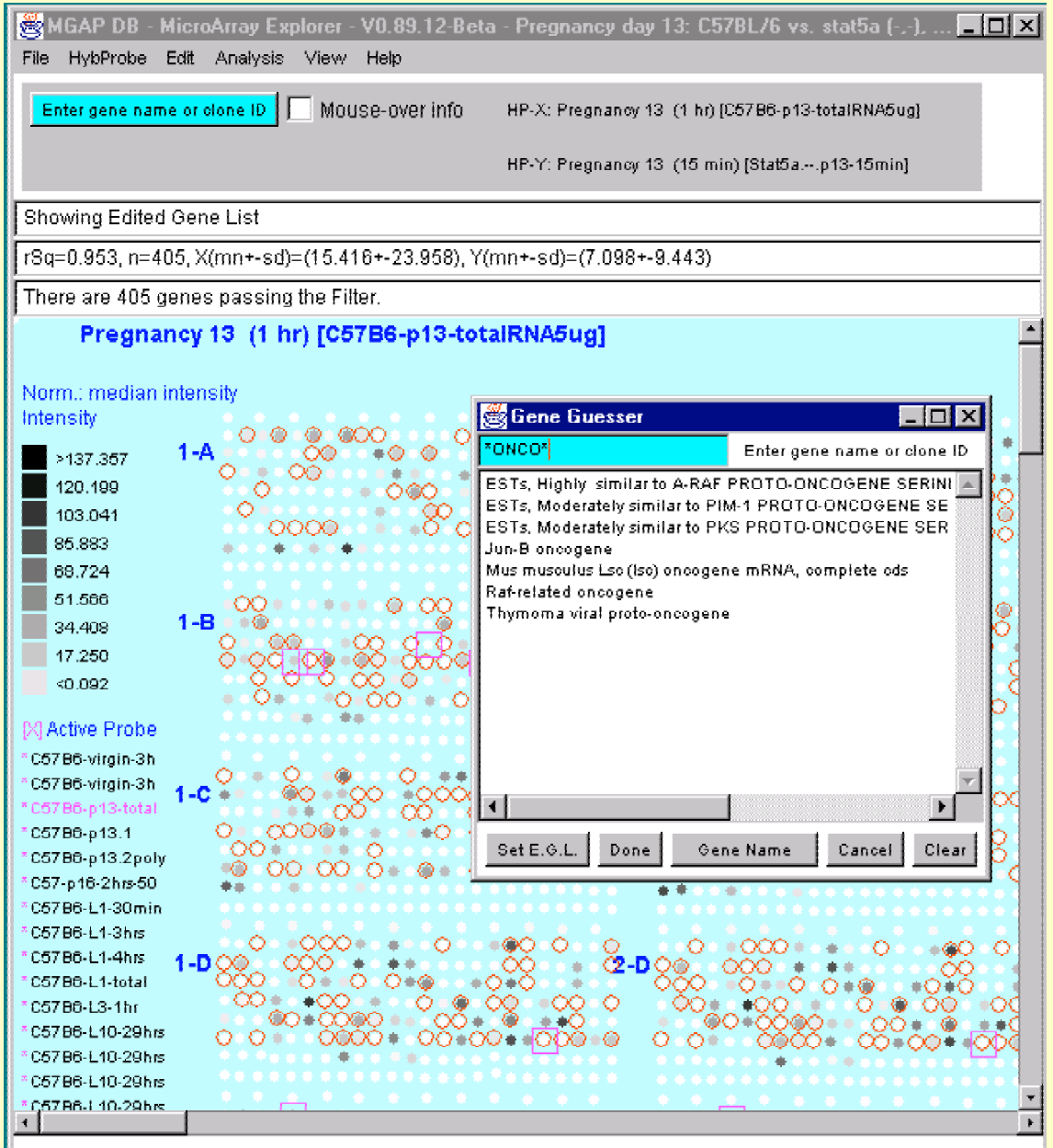
II.2 Opening a Database from Local Disk

- In stand-alone mode, you may browse a project database containing startup databases.

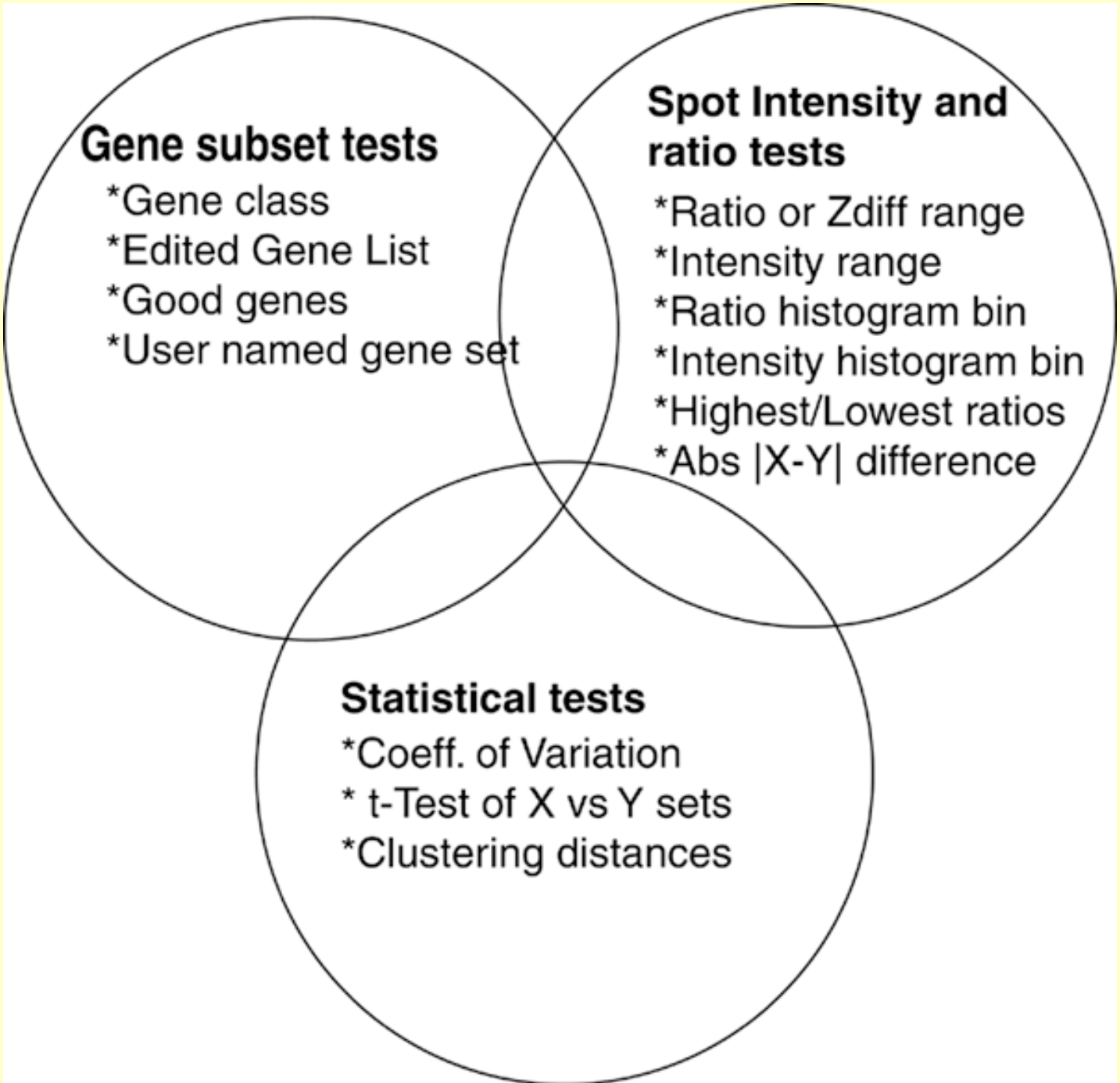


II.3 Specify Gene or Gene Subset by Name

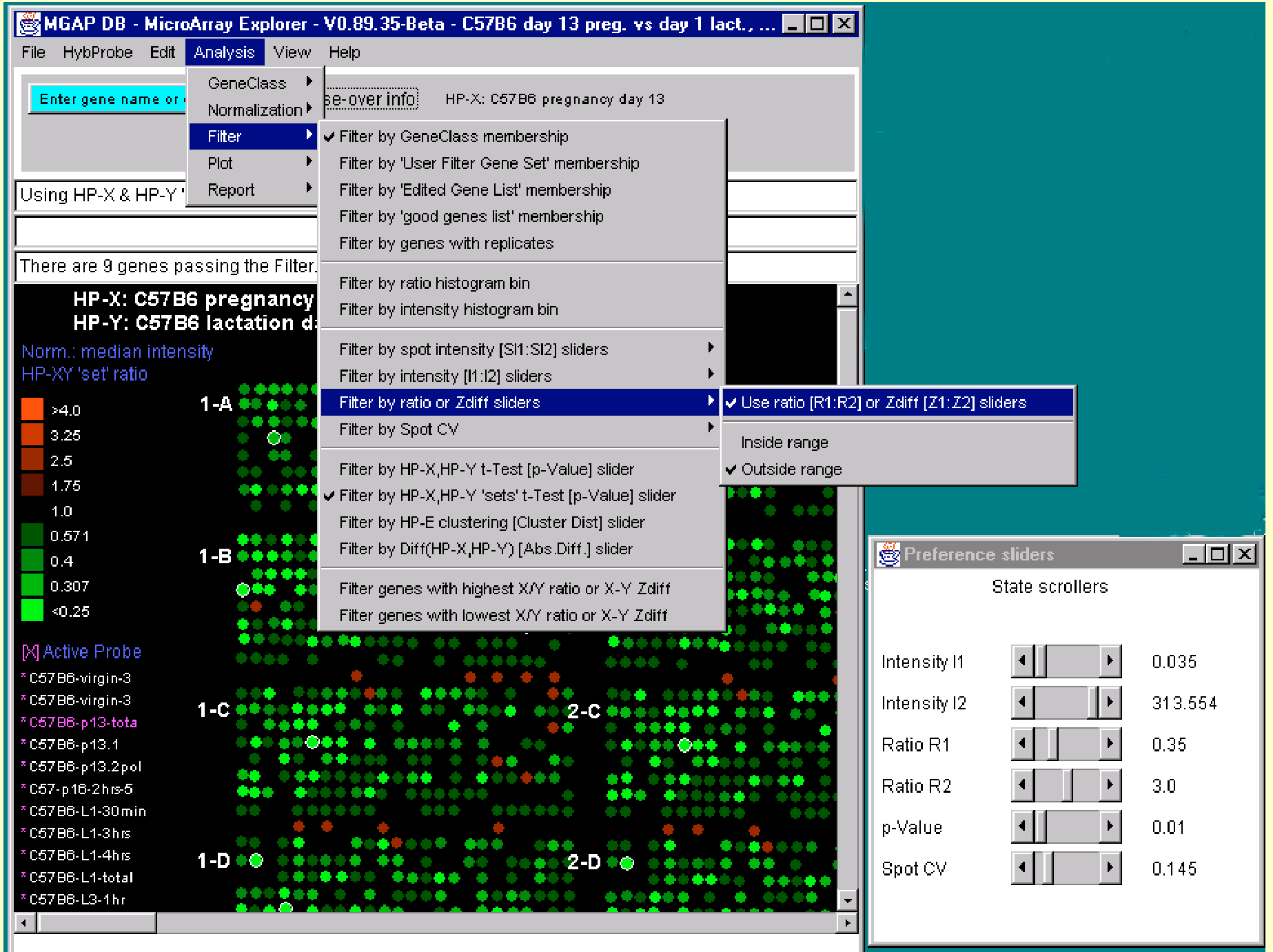
- Specify gene or gene subset by gene name guesser using wildcard sub-strings eg. “*ONCO*” indicated by magenta boxes in pseudo image - saved in ‘Edited Gene List’.



II.4 Gene Data Filter is Intersection of Tests



II.5 MAExplorer Data Filter Menu



II.6a Gene Set Operations Help Manage Data and Search Results

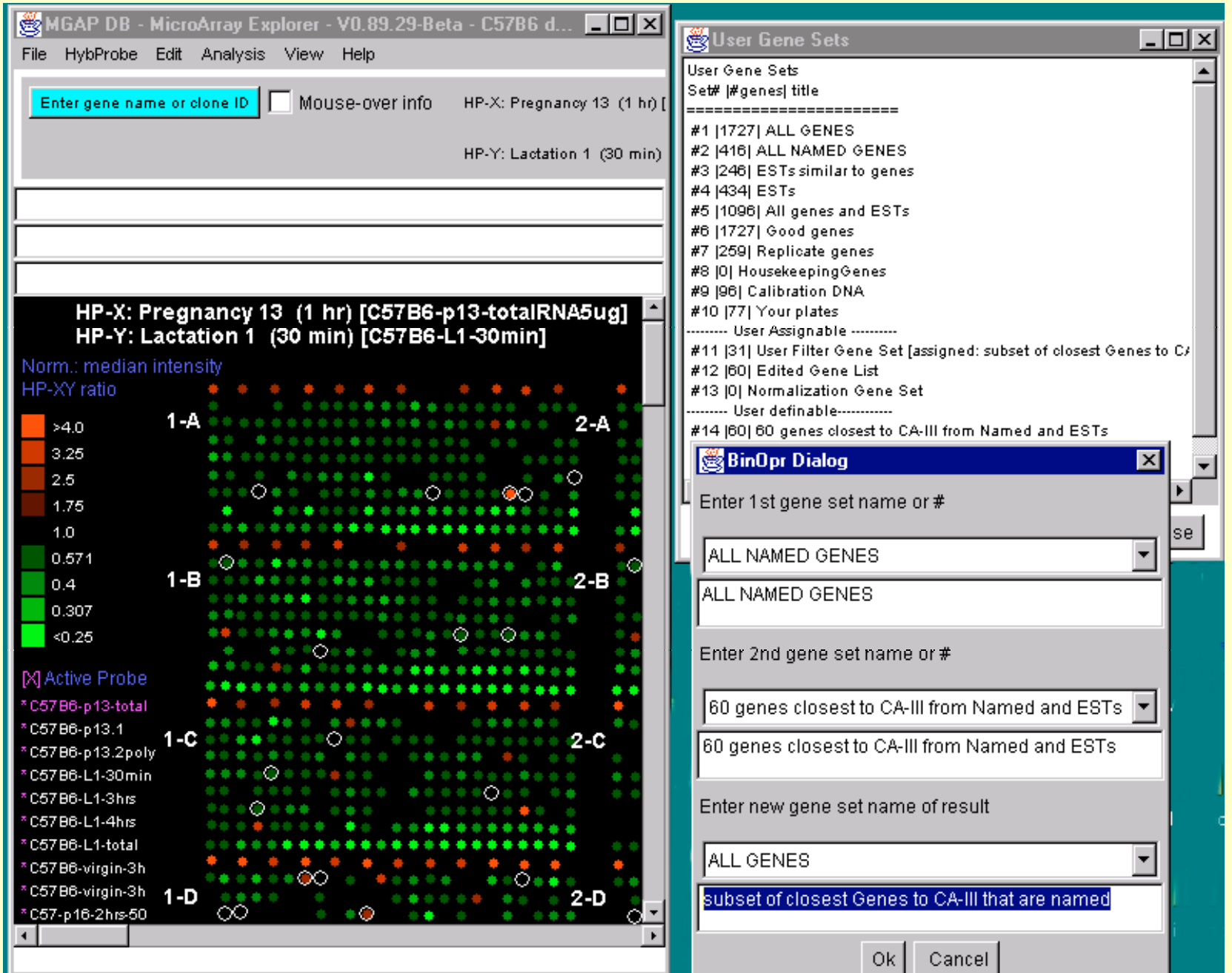
- All gene sets are named with a directory of existing sets

- Set operations (AND, OR, DIFFERENCE) may be used to create new derived sets

- Special sets:
  1. Filtered genes set holds genes passing the data filter
  2. Edited Gene List holds results of clustering or editing
  3. Normalization set may be used as normalization method
  4. User data filter set may be used as a data filter

- Sets are saved when the session is saved, restored when MAExplorer is restarted

II.6b Gene Set Operations - e.g. 'AND' of Two Sets

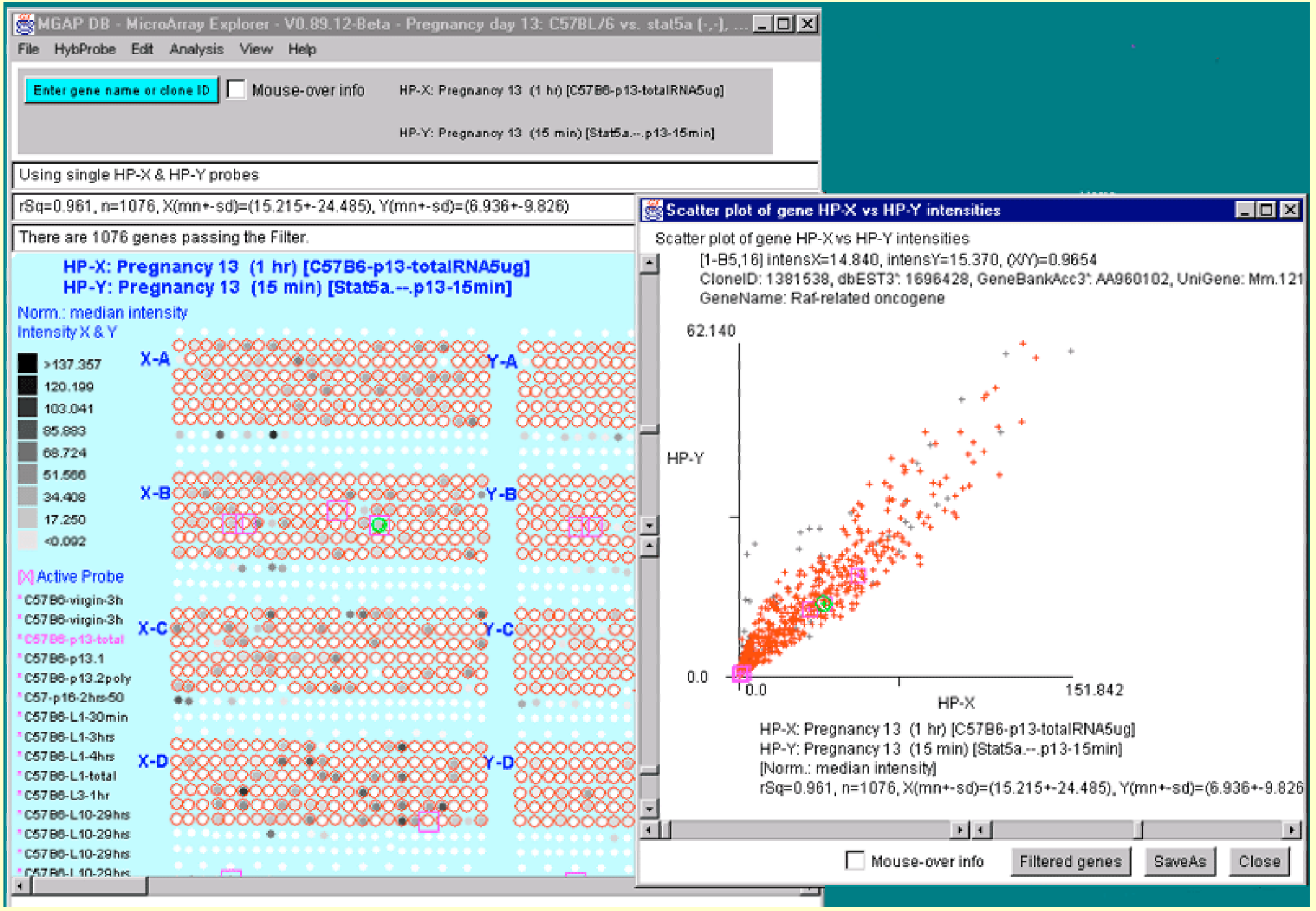


II.7 Summary of Types of Plots

- Plots allow visualization and direct manipulation of gene data
- 1. Pseudo array image - intensity, ratio (X/Y)
- 2. Zoomable scatter plots - X vs Y, Cy3 vs Cy5, duplicate spots
- 3. Histograms - ratio and intensity
- 4. Expression profiles - individual genes and overlay plots
- 5. Silhouette plots - similarity clusters, K-means clustering
- 6. Hierarchical clustering - clustergram, dendrogram

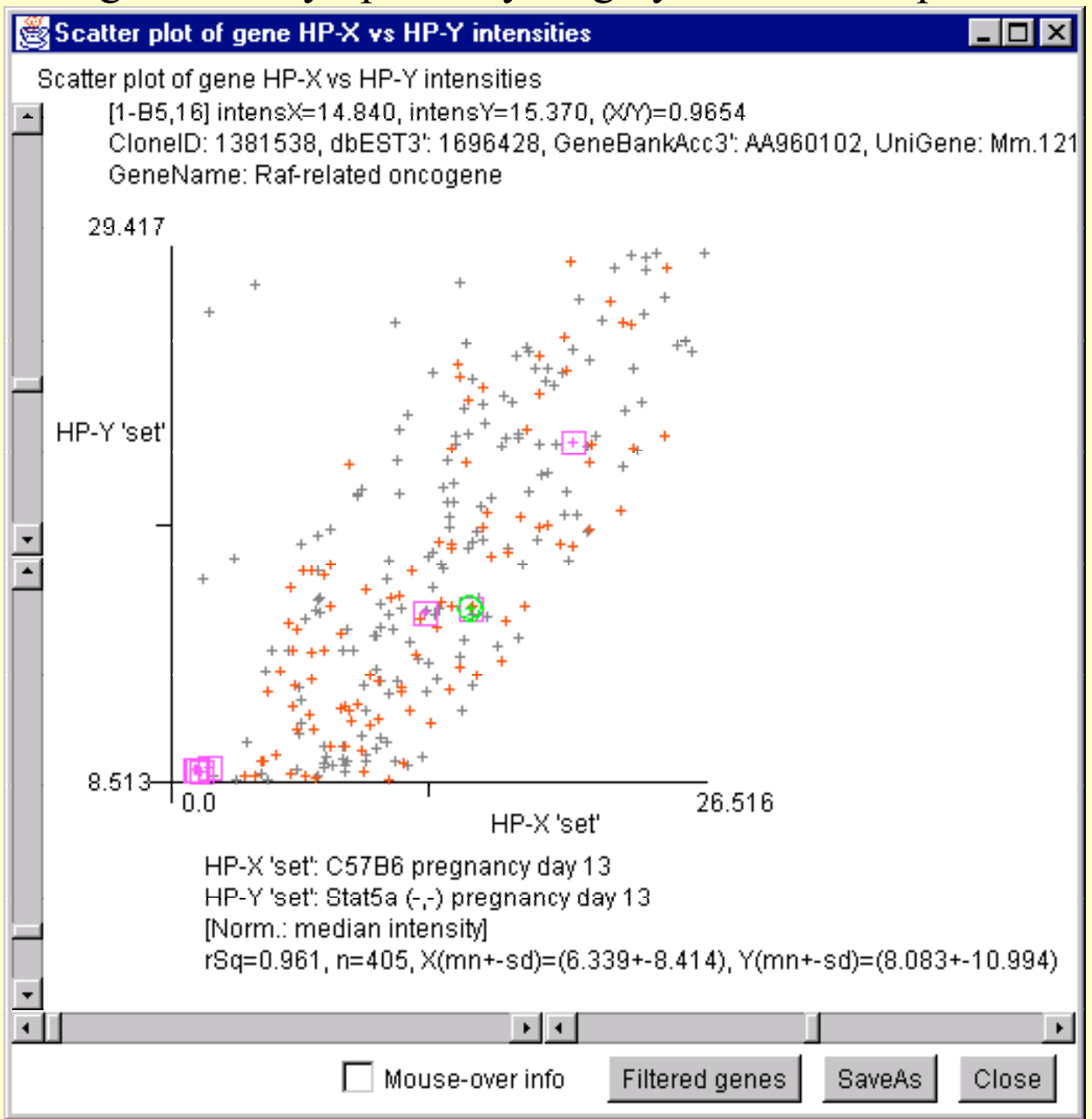
II.8a Scatter Plots of Two Conditions

- X-Y scatter plot of two 13-day pregnancy samples: C57B6 vs Stat5a (-,-) [MGAP]

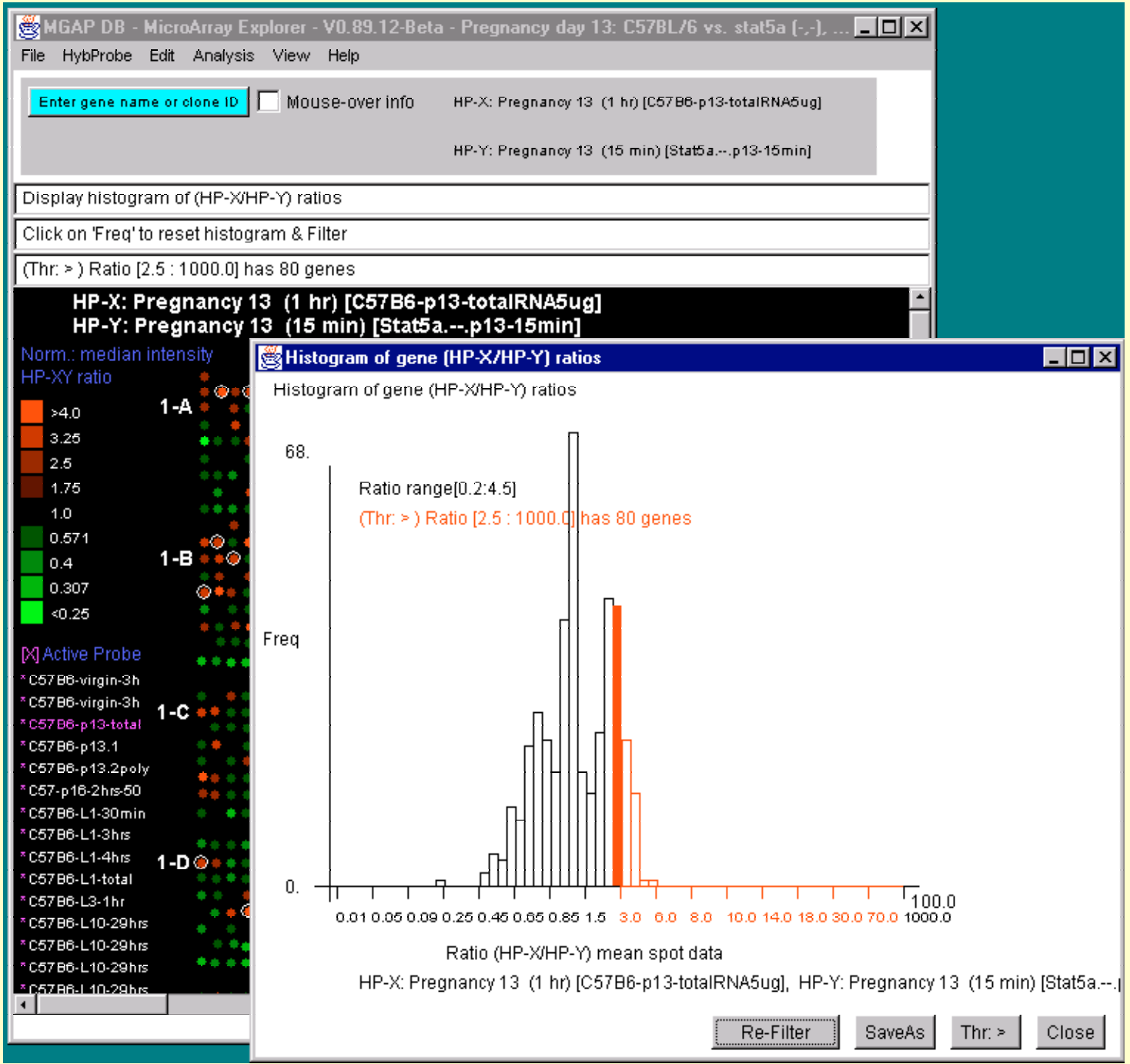


II.8b Zoom X-Y Scatter Plot by Scrollers

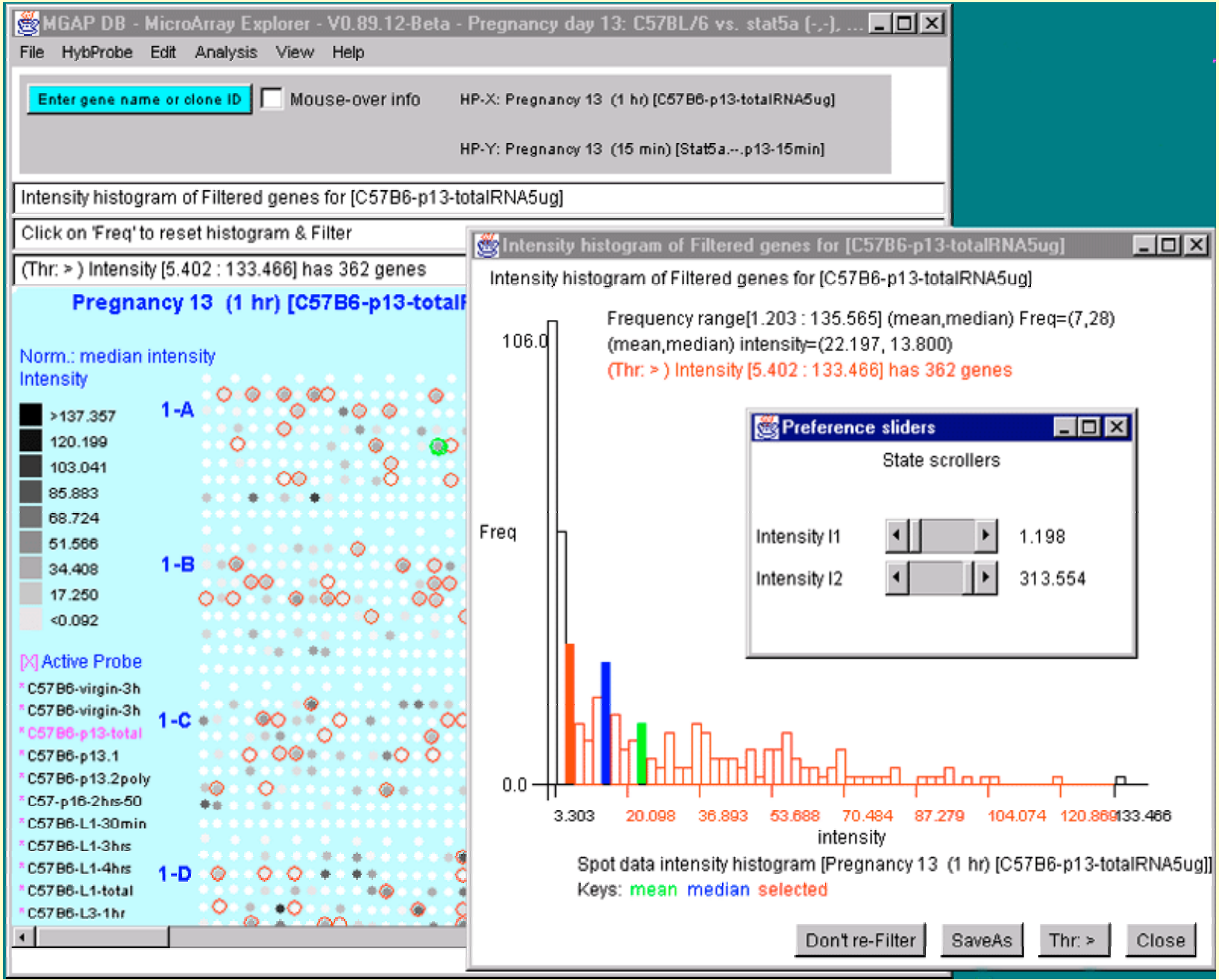
- Zoomed-in on Raf-related oncogene (green circle) using scrollbars
- Genes not passing Filter may optionally be grayed out in the plot



II.9a Filter Genes by X/Y Ratio-Histogram Bin(s)



II.9b Filter Genes by Intensity-Histogram Bin(s)



II.10 Types of Reports

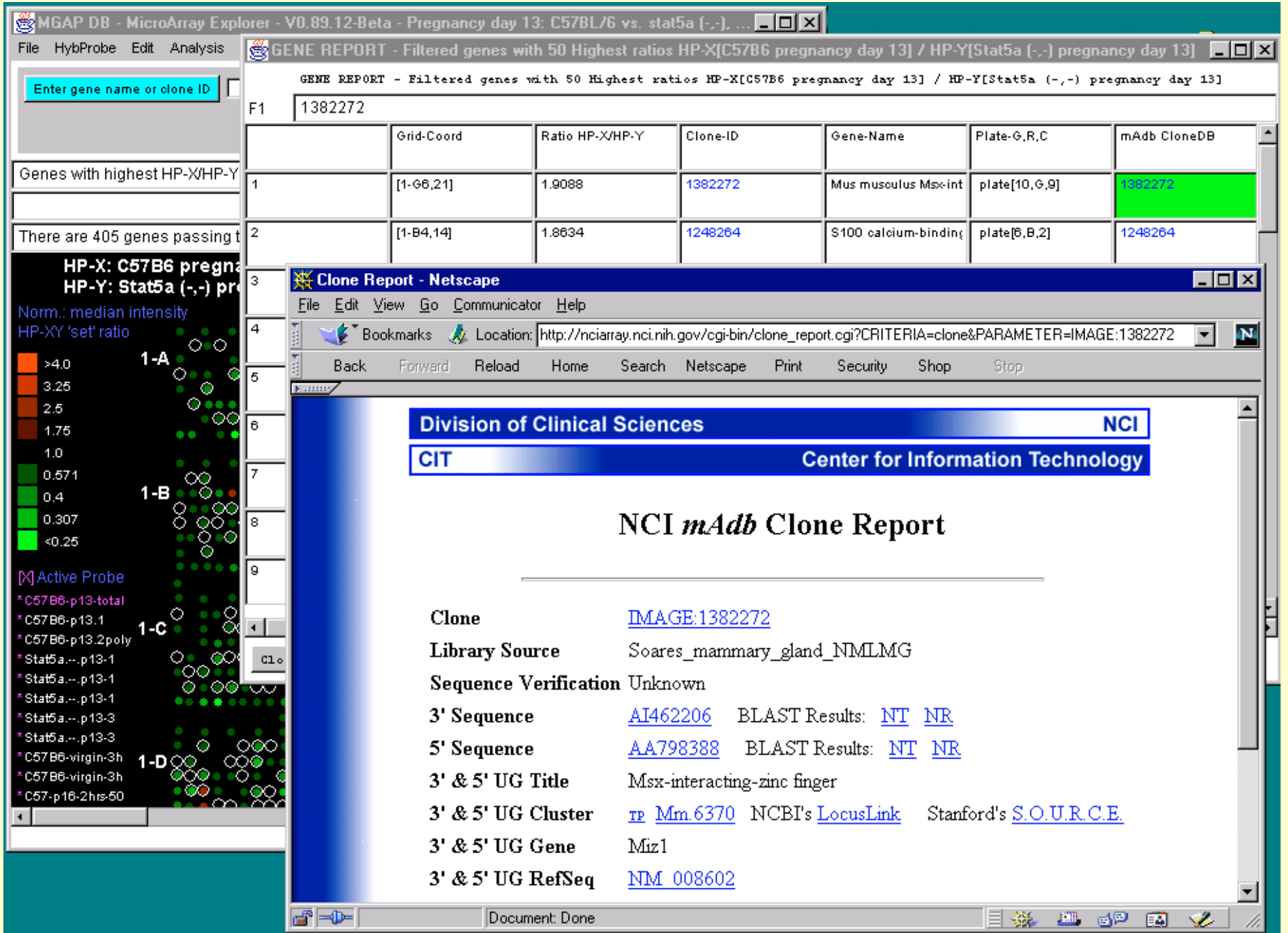
- Data reported as:
    1. Web-accessible scrollable dynamic spreadsheets or
    2. tab-delimited text exportable to Excel

- Gene set reports - linked to UniGene, GenBank, mAdb, etc.

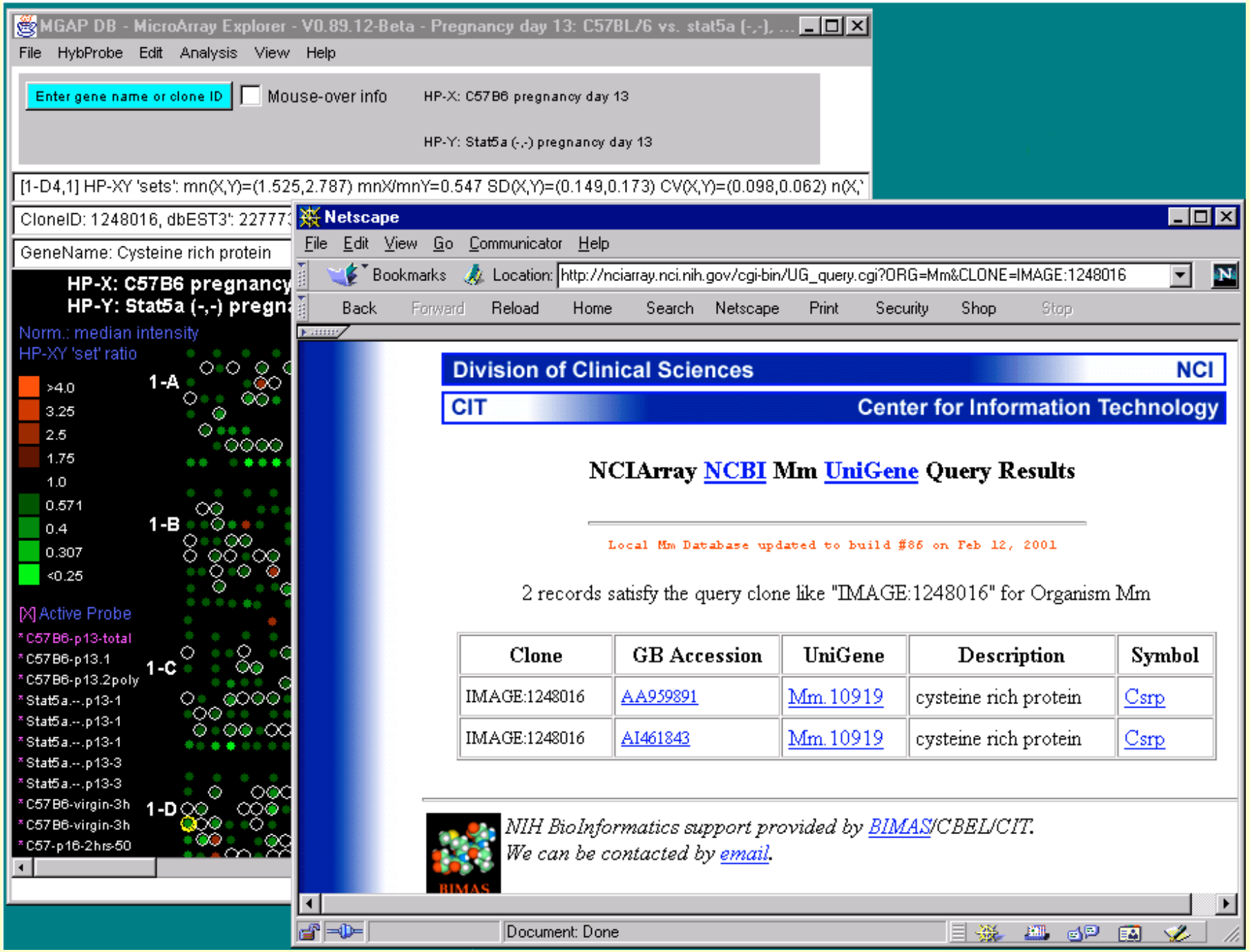
- Array sample reports - linked to histology and model Web pages

- Pop-up Web browser on specific data from dynamic reports or plots

II.11a Scrollable Dynamic Gene Reports - mAdb

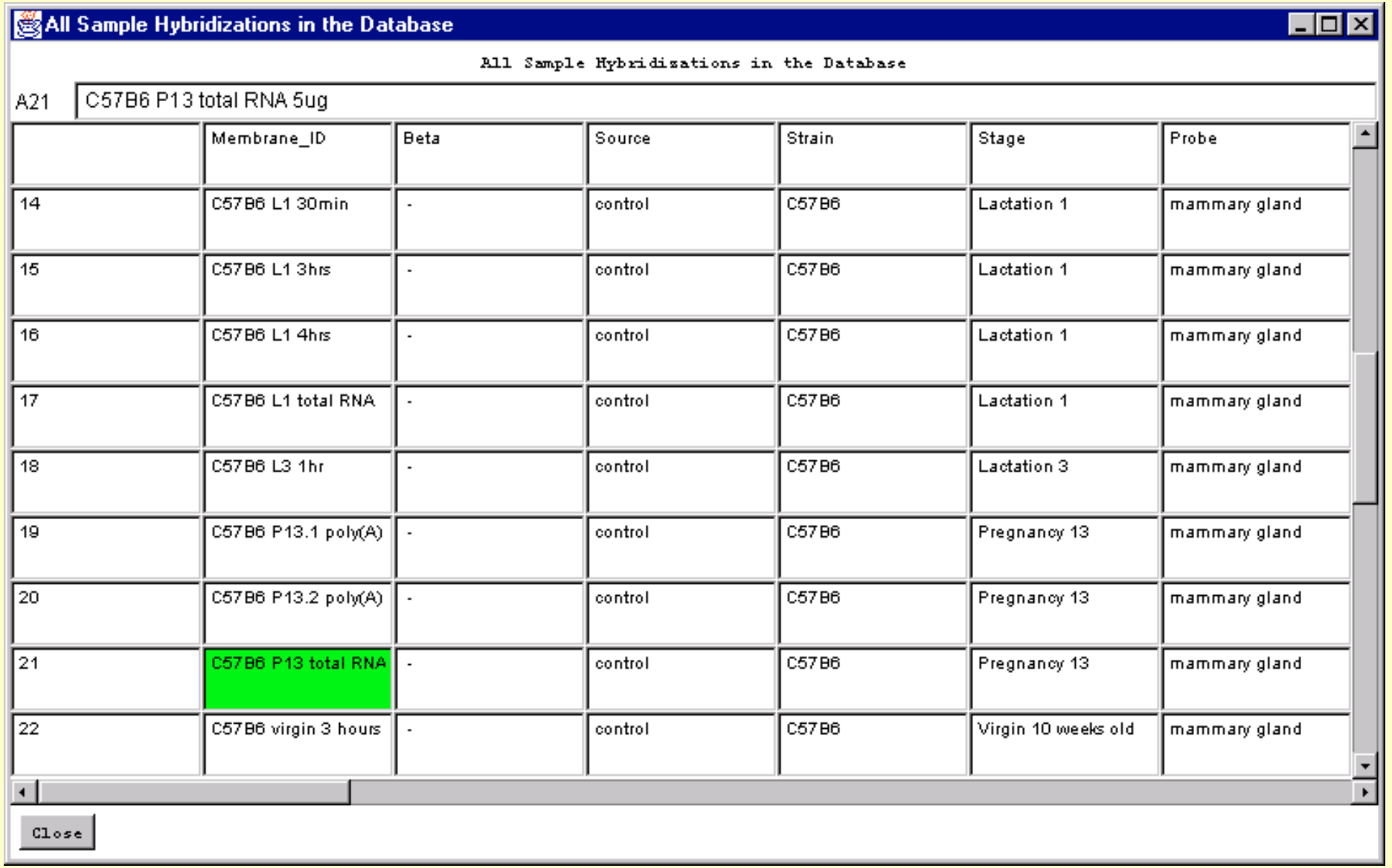


II.11b Scrollable Dynamic Gene Report - UniGene



II.12a Sample Information Array Reports

- Details are available on all hybridized array samples



	Membrane_ID	Beta	Source	Strain	Stage	Probe
14	C57B6 L1 30min	-	control	C57B6	Lactation 1	mammary gland
15	C57B6 L1 3hrs	-	control	C57B6	Lactation 1	mammary gland
16	C57B6 L1 4hrs	-	control	C57B6	Lactation 1	mammary gland
17	C57B6 L1 total RNA	-	control	C57B6	Lactation 1	mammary gland
18	C57B6 L3 1hr	-	control	C57B6	Lactation 3	mammary gland
19	C57B6 P13.1 poly(A)	-	control	C57B6	Pregnancy 13	mammary gland
20	C57B6 P13.2 poly(A)	-	control	C57B6	Pregnancy 13	mammary gland
21	C57B6 P13 total RNA	-	control	C57B6	Pregnancy 13	mammary gland
22	C57B6 virgin 3 hours	-	control	C57B6	Virgin 10 weeks old	mammary gland

II.12b Correlation Report of All Samples

- Sample vs Sample correlation coefficient report for *current set* of **Filtered** genes

HP vs. HP correlation coefficients table, Pregnancy 13 days: C57BL/6 vs. stat5a (-,-), 8 probes

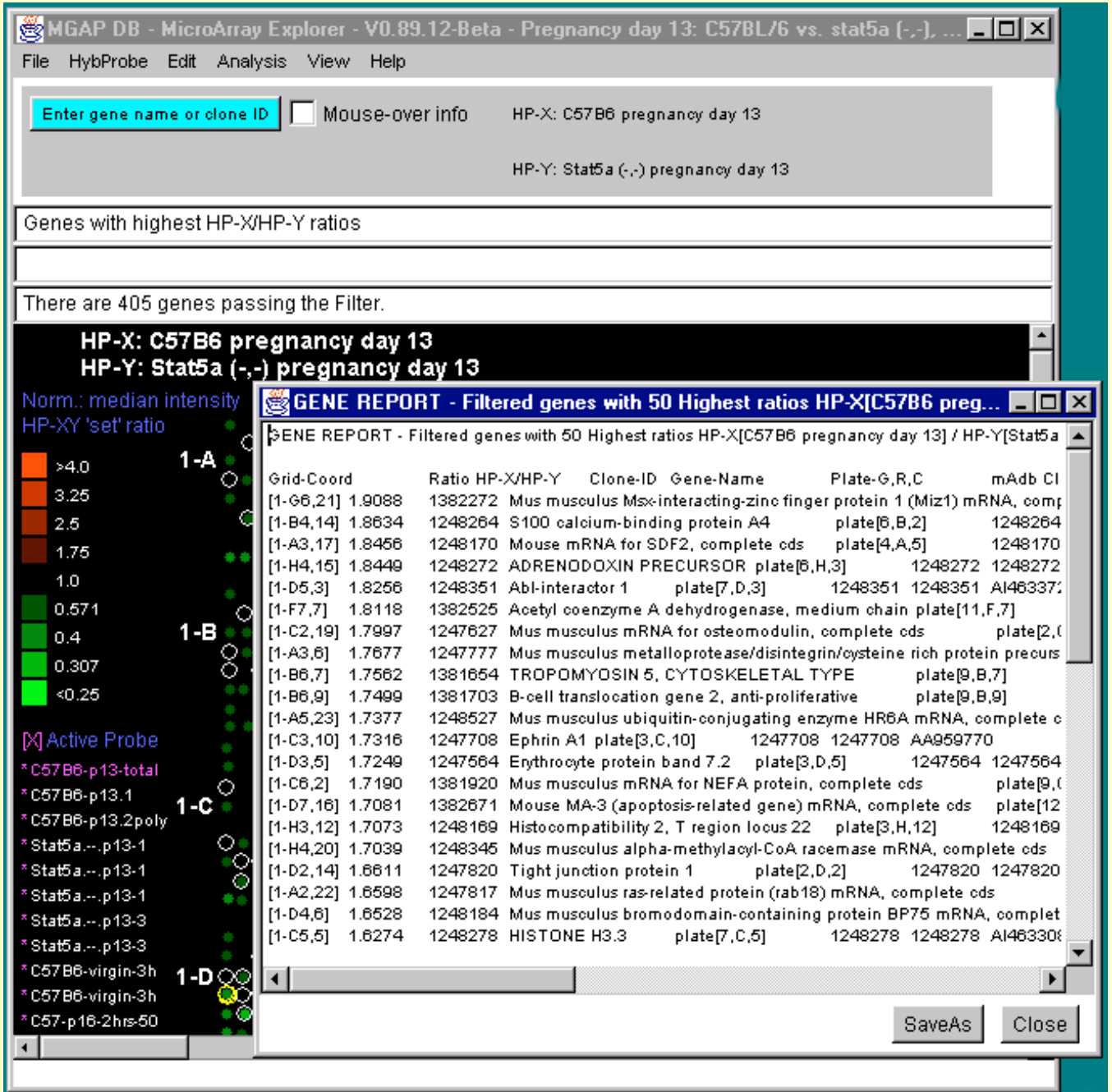
HP vs. HP correlation coefficients table, Pregnancy 13 days: C57BL/6 vs. stat5a (-,-), 8 probes

D2 rSq=0.982, n=405, HP:2(mn+-sd)=(1+-0), HP:3(mn+-sd)=(1+-1)

		C57B6-p13-totalRNA	C57B6-p13.1	C57B6-p13.2poly-A	Stat5a.--.p13-15min	Stat5a.--.p13-15min2
1	C57B6-p13-totalRNA	-	rSq=0.715, n=405, HF	rSq=0.729, n=405, HF	rSq=0.953, n=405, HF	rSq=0.958, n=405, HF
2	C57B6-p13.1	-	-	rSq=0.982, n=405, HF	rSq=0.756, n=405, HF	rSq=0.757, n=405, HF
3	C57B6-p13.2poly-A	-	-	-	rSq=0.772, n=405, HF	rSq=0.773, n=405, HF
4	Stat5a.--.p13-15min	-	-	-	-	rSq=0.997, n=405, HF
5	Stat5a.--.p13-15min2	-	-	-	-	-
6	Stat5a.--.p13-1hr2	-	-	-	-	-
7	Stat5a.--.p13-30min	-	-	-	-	-
8	Stat5a.--.p13-30min2	-	-	-	-	-
9	-	-	-	-	-	-

Close

II.13 Tab-Delimited Reports Exportable to Excel

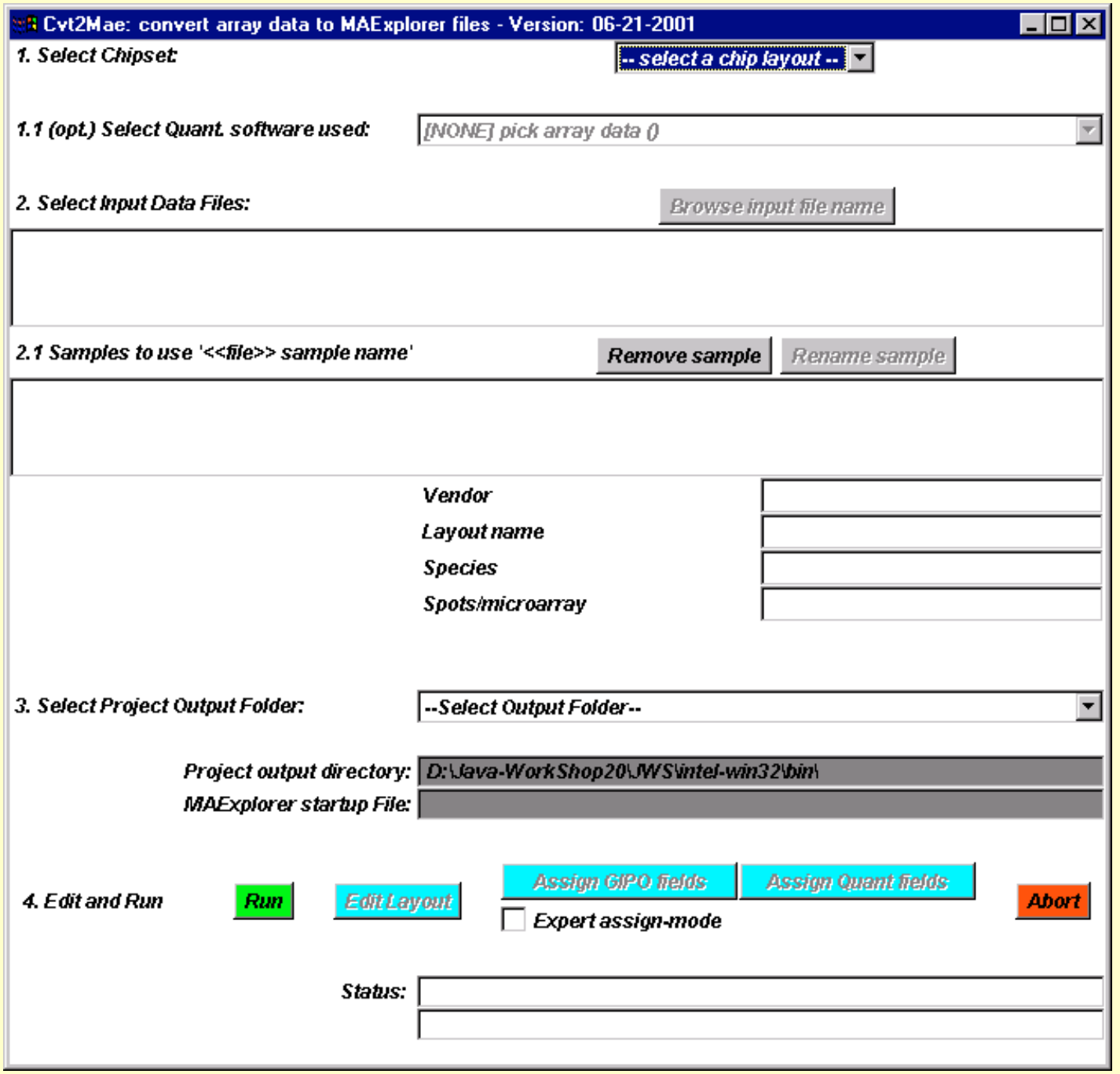


II.14 [Undergoing *Beta*-testing] Data Conversion for MAExplorer: Cvt2Mae

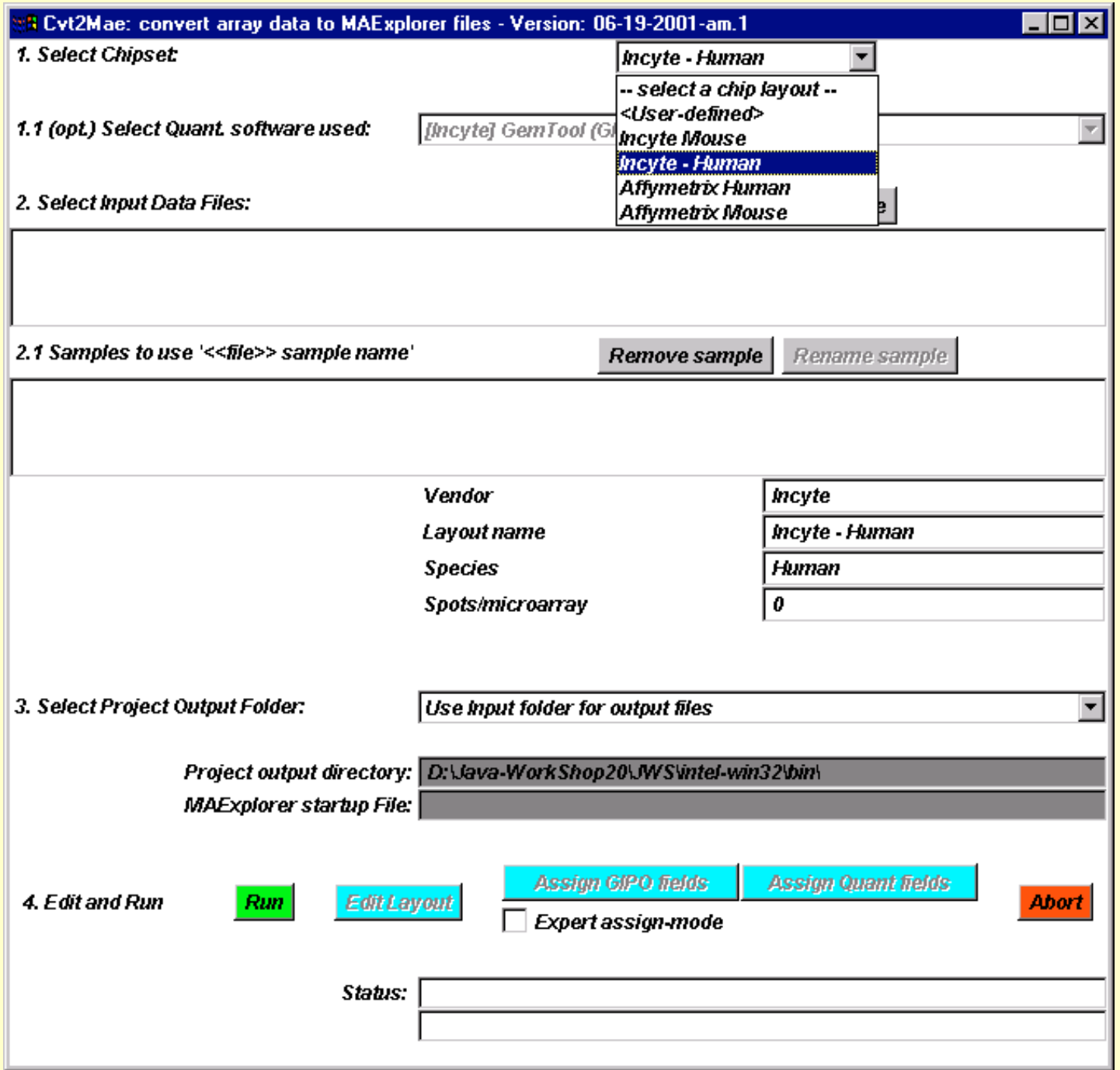
- Cvt2Mae reads and converts a variety of array data to MAExplorer format:
  1. commercial arrays (Incyte, Affymetrix, etc.)
  2. user-defined academic arrays

- Cvt2Mae is undergoing testing and will be made available for download through the MAExplorer Web site

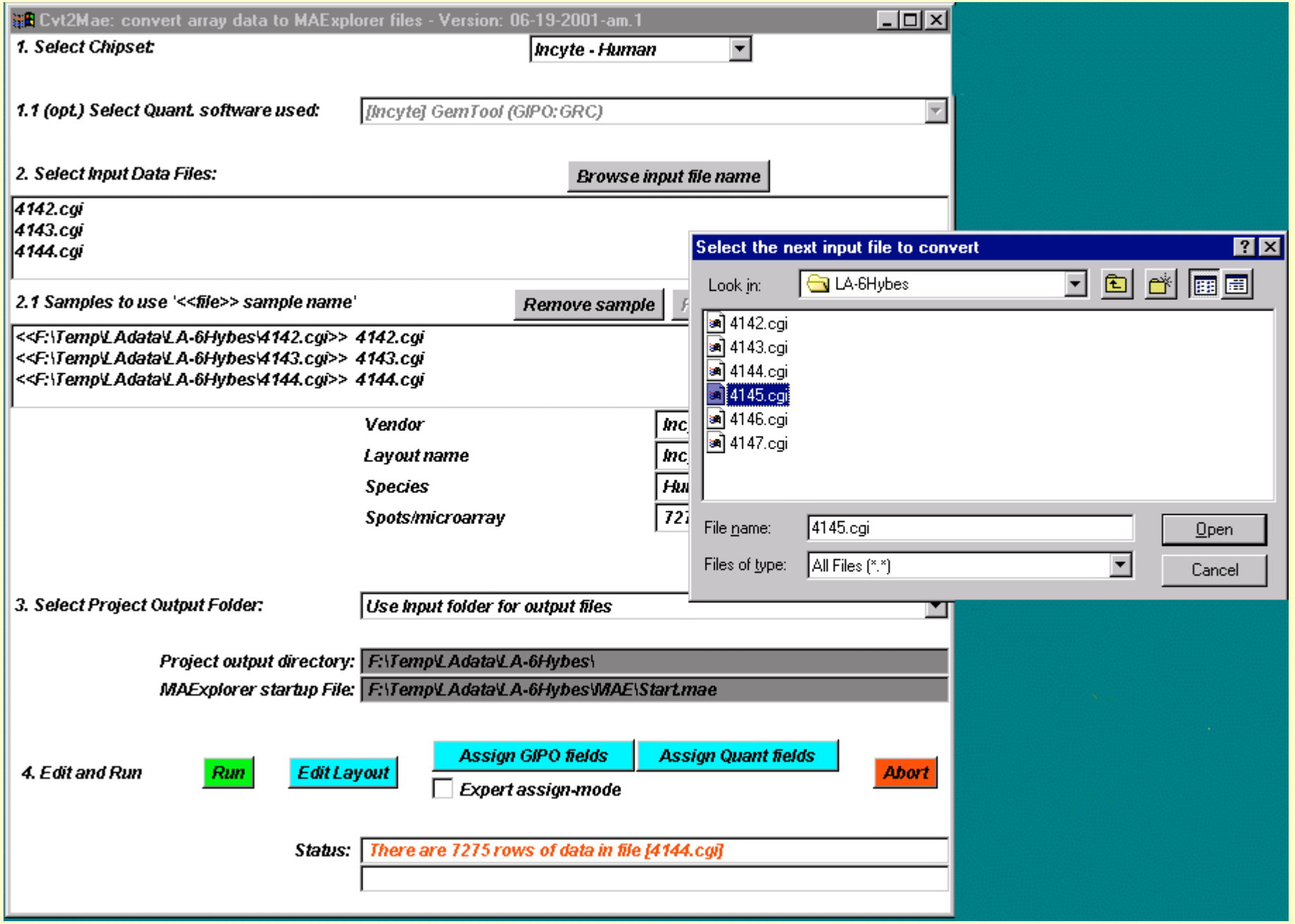
II.15a Initial State of Cvt2Mae Program



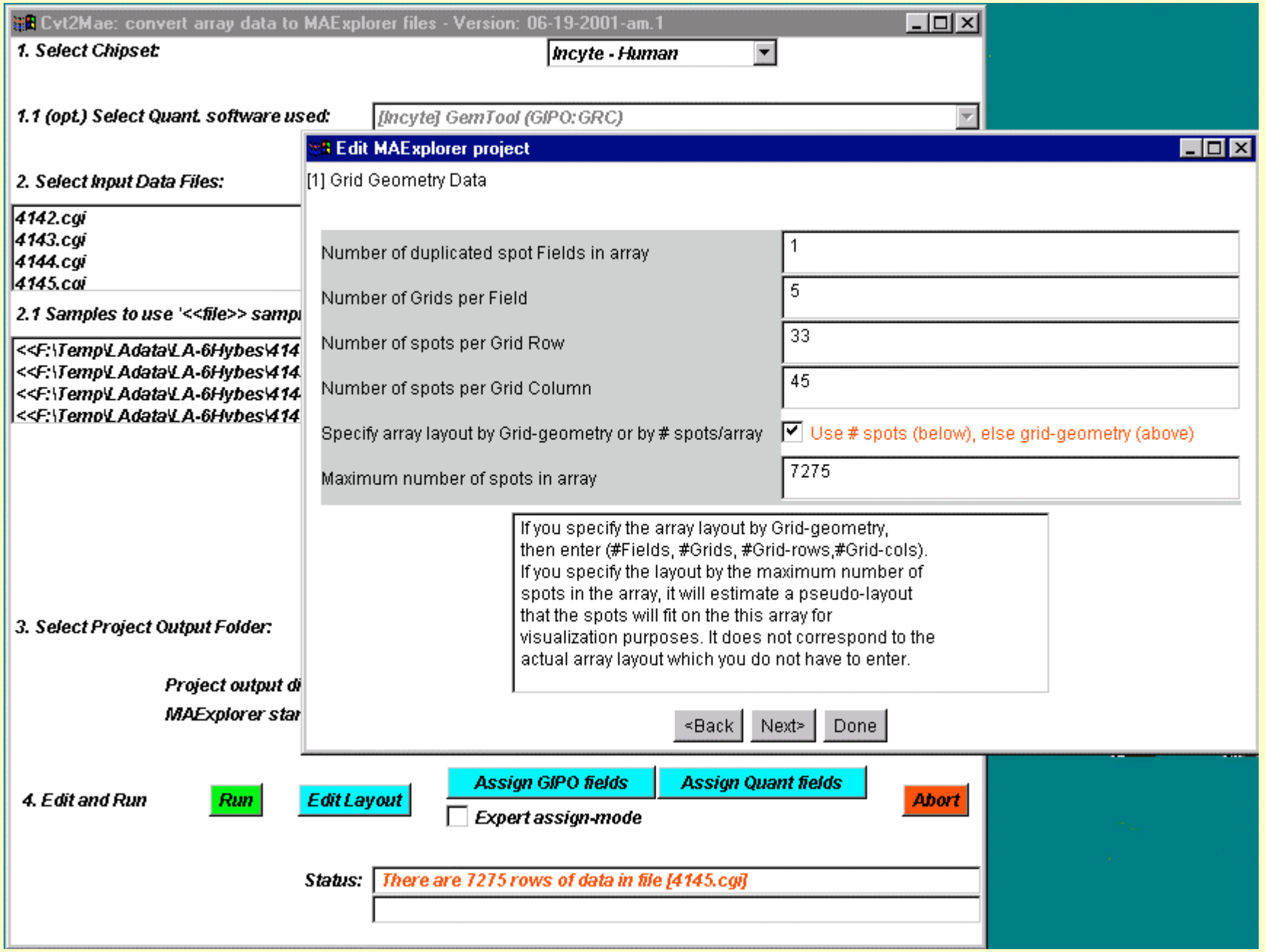
II.15b Select a Chip Array-Layout



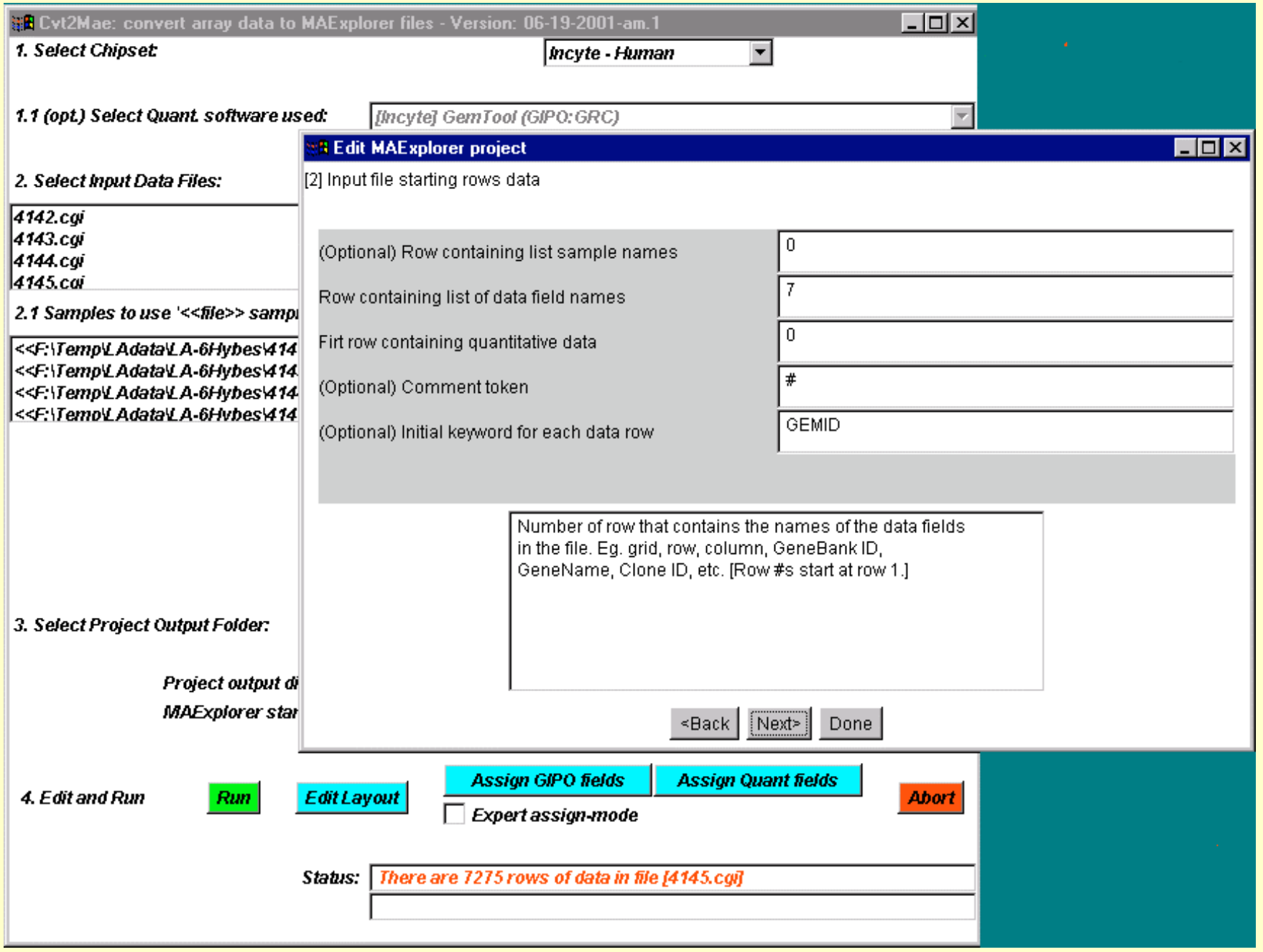
II.16c Select Input Files with Input-Browser



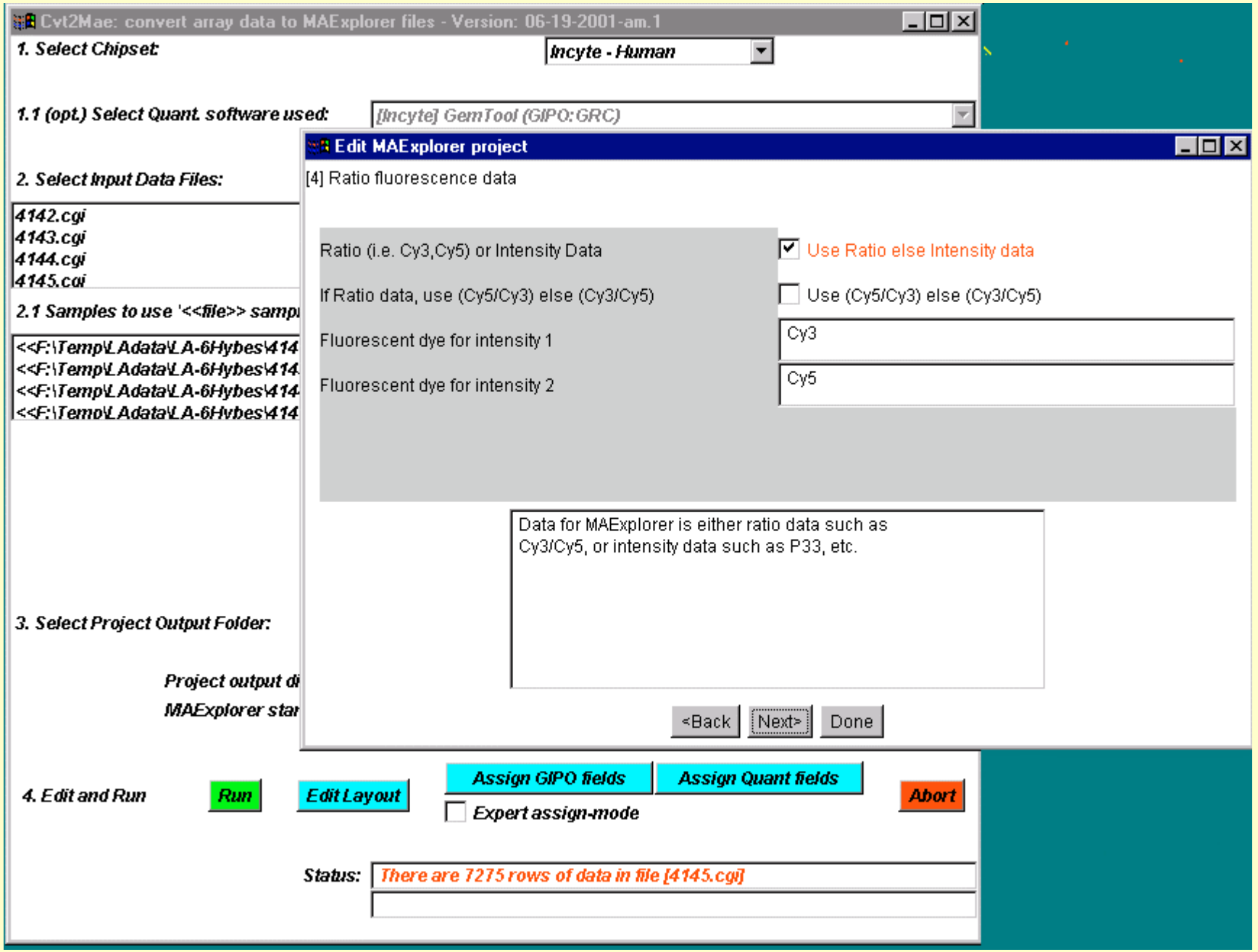
II.17d Optional Edit Layout - Defining Geometry



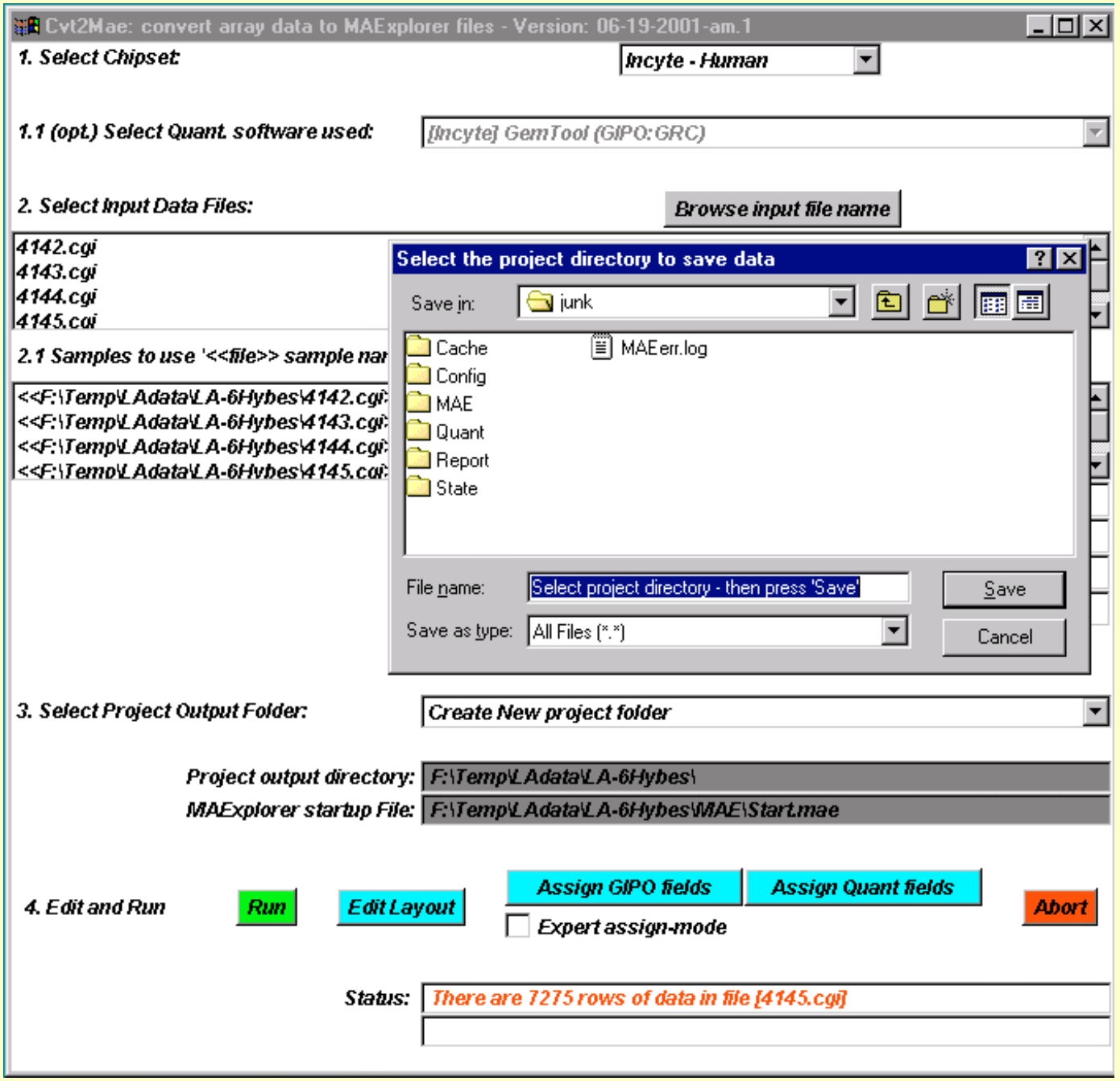
II.17e Edit Layout - Specify Input File Rows



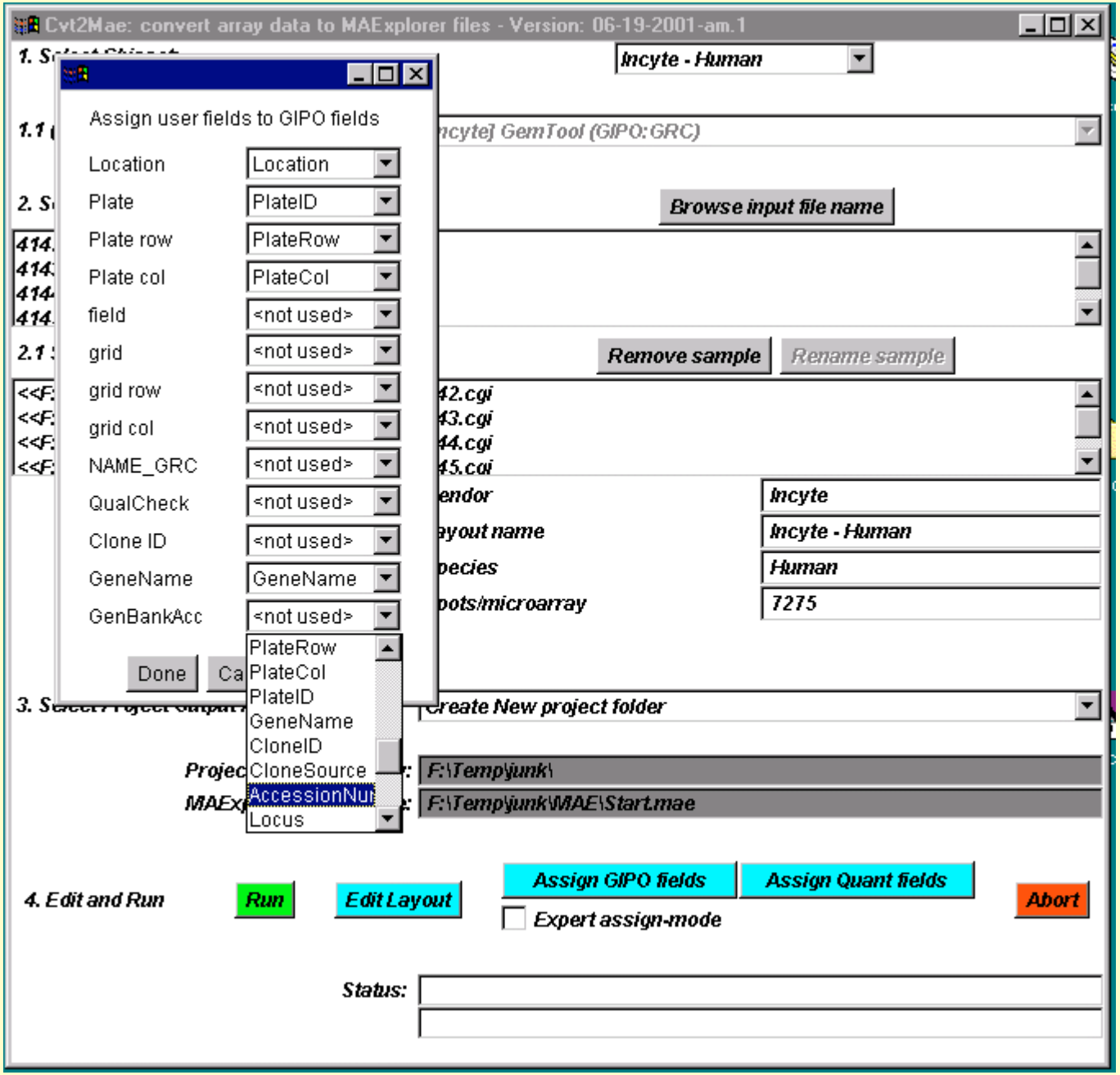
II.17f Edit Layout - Intensity or Ratio Data



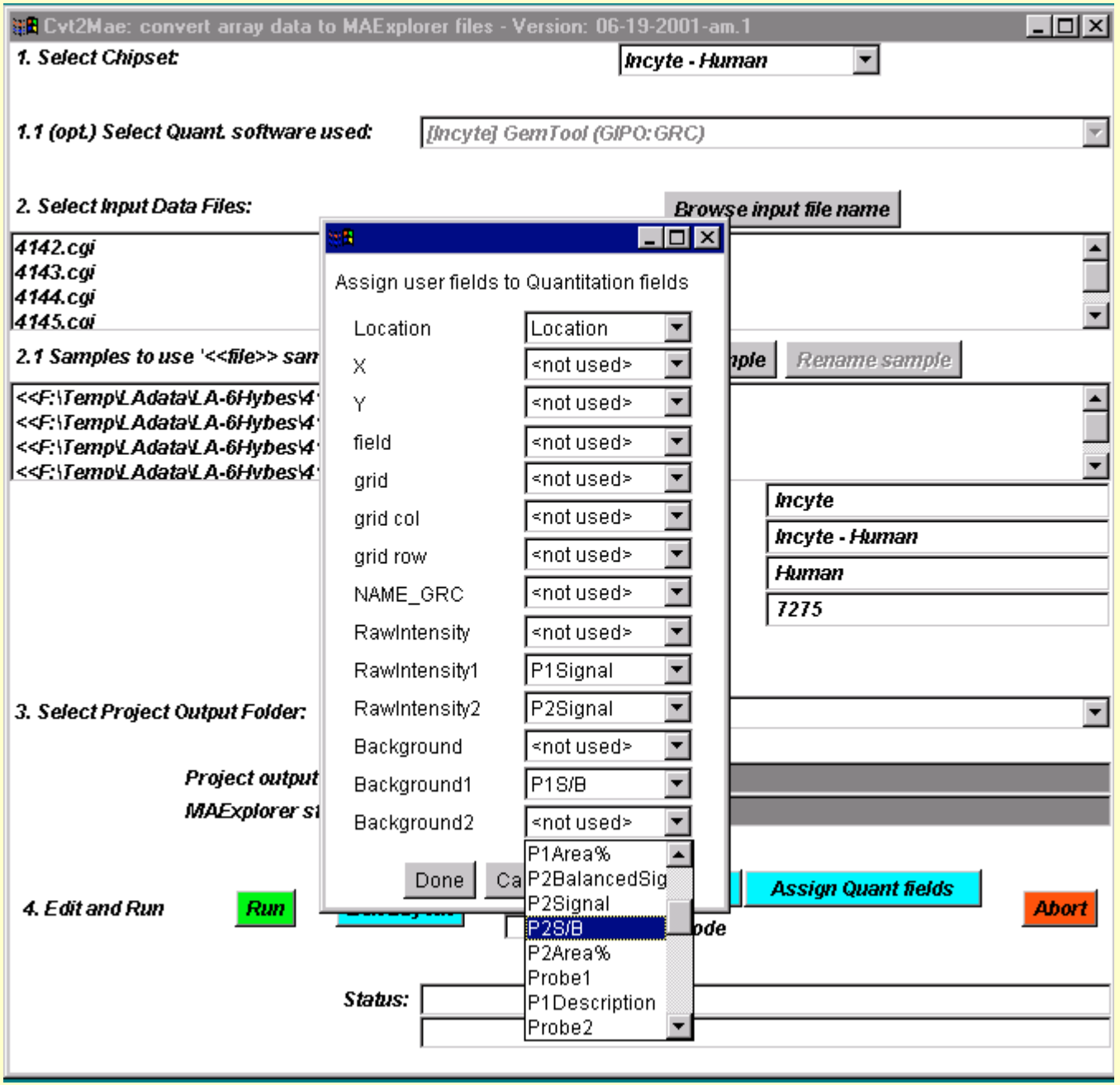
II.17g Specify Output Folder for Converted Data



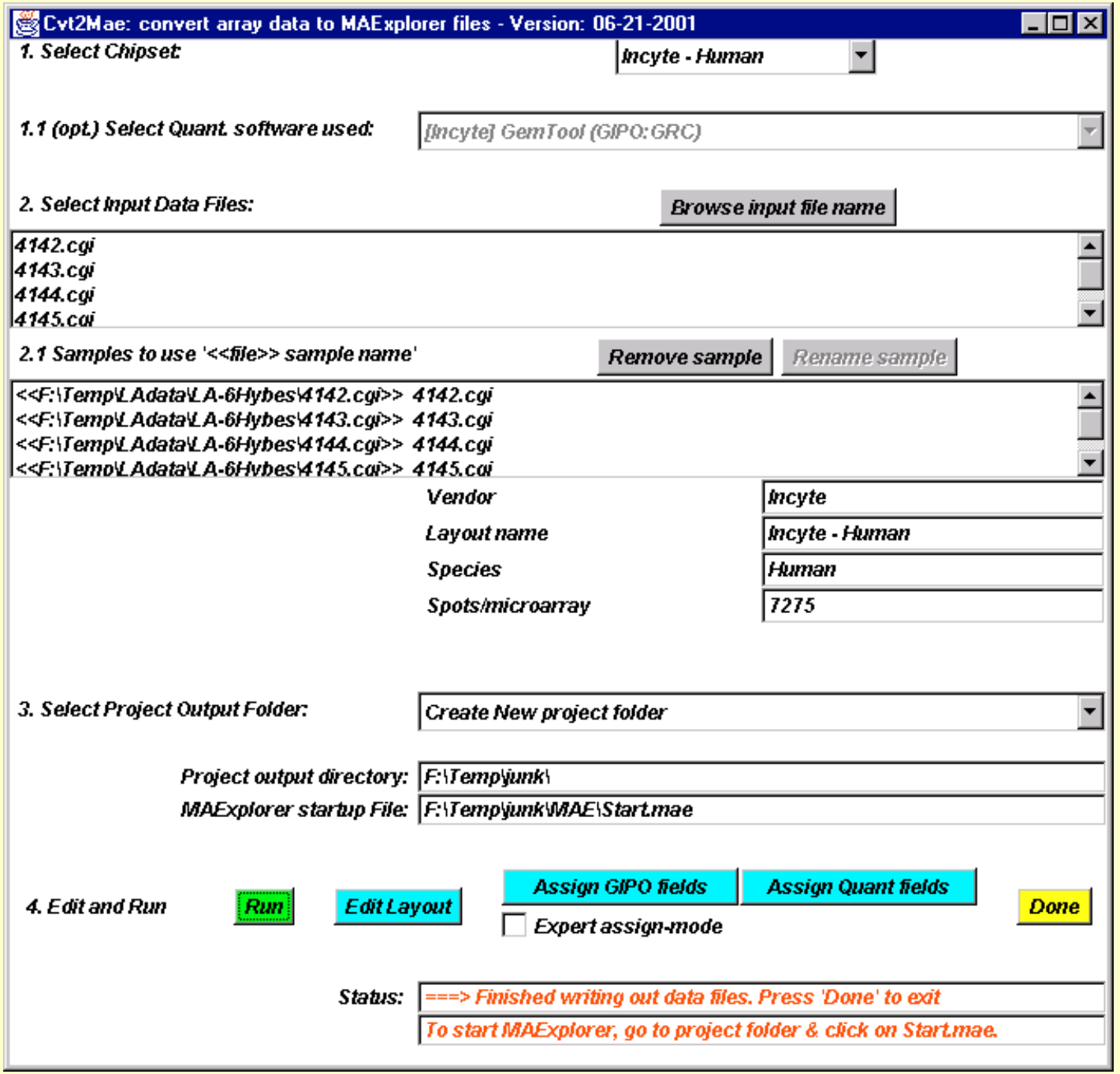
II.17h Optionally Assign GIPO Fields



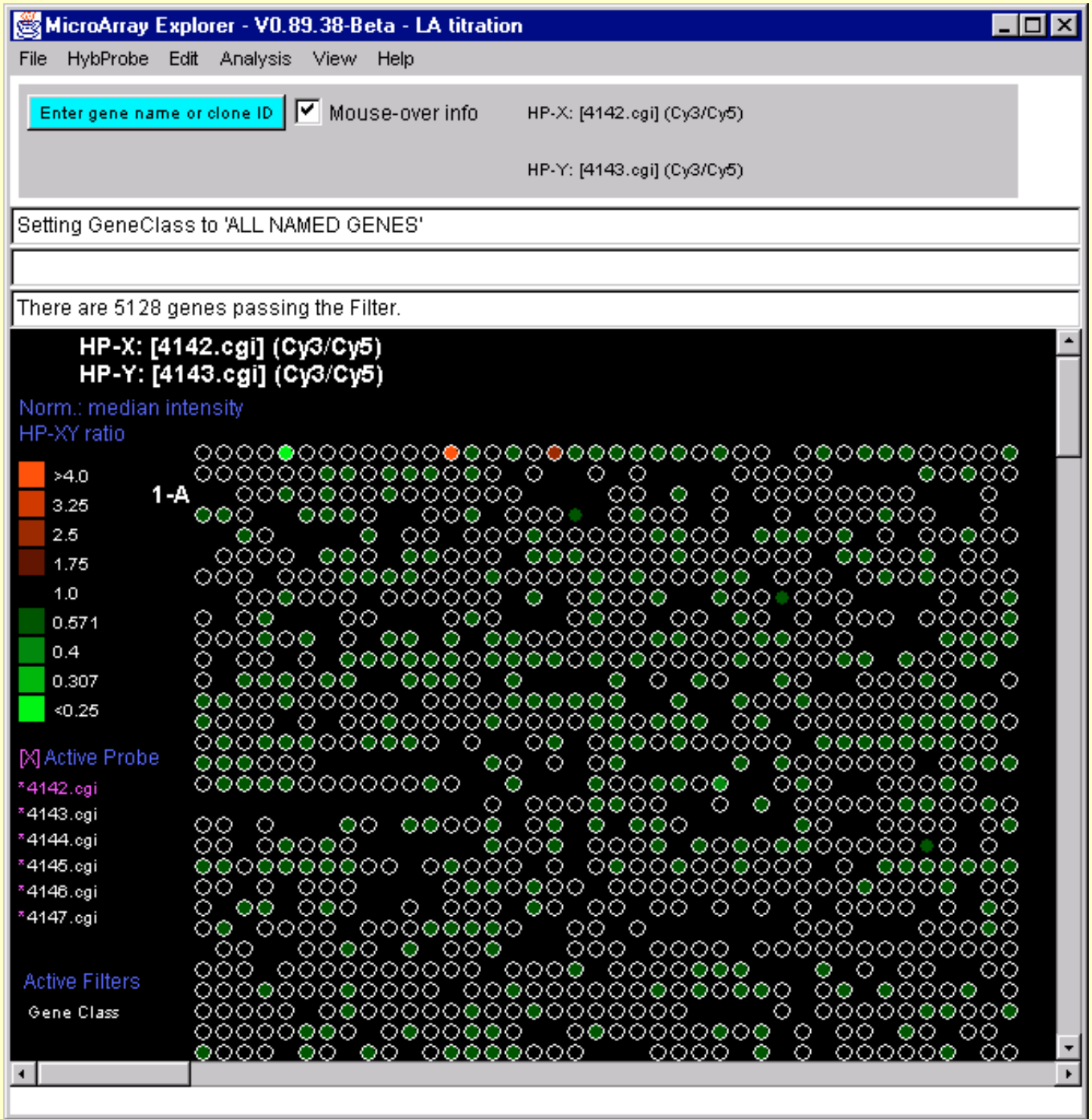
II.17i Optionally Assign Quantified Data Fields



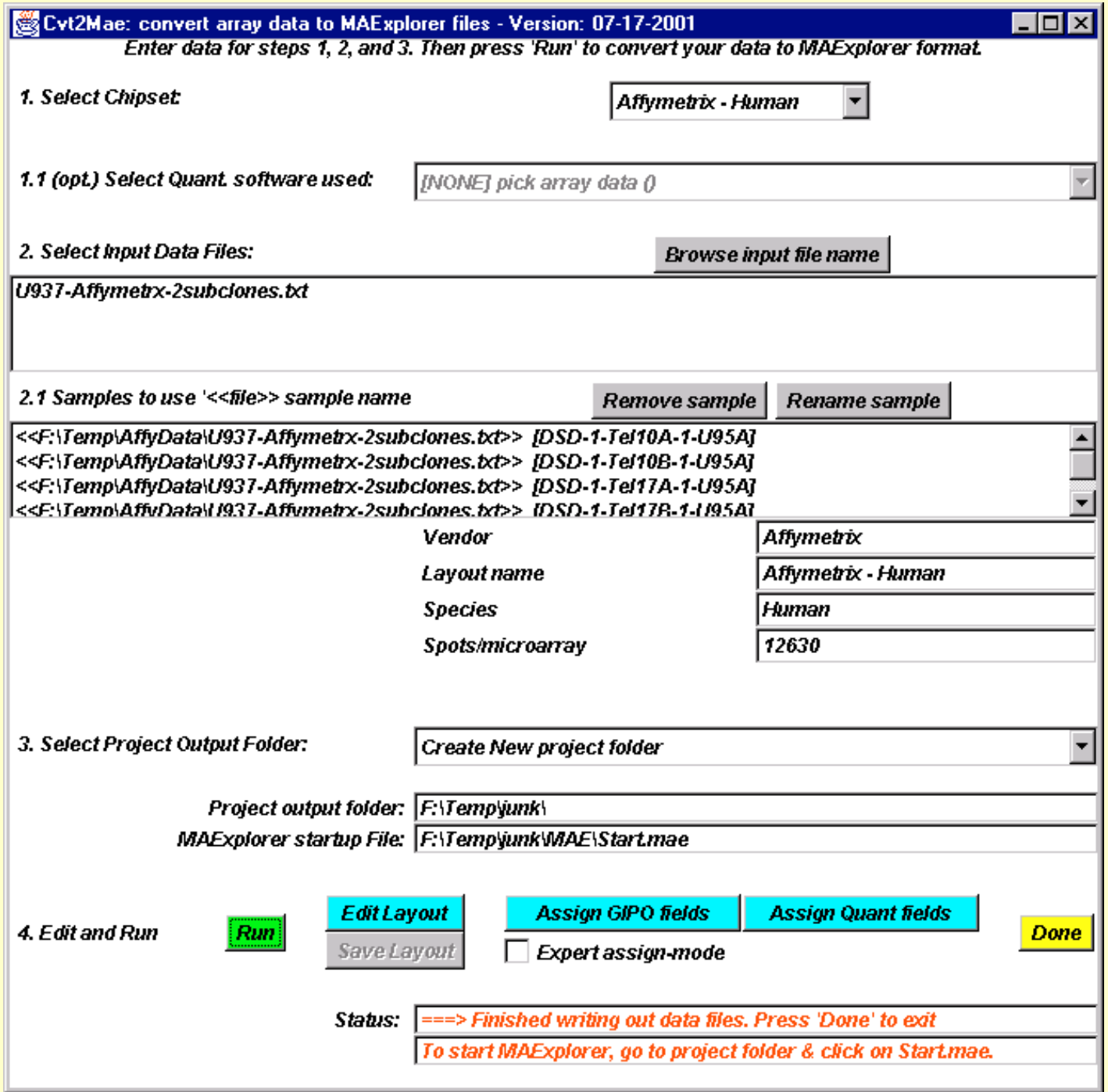
II.18a Converted Incyte files Ready to Analyze



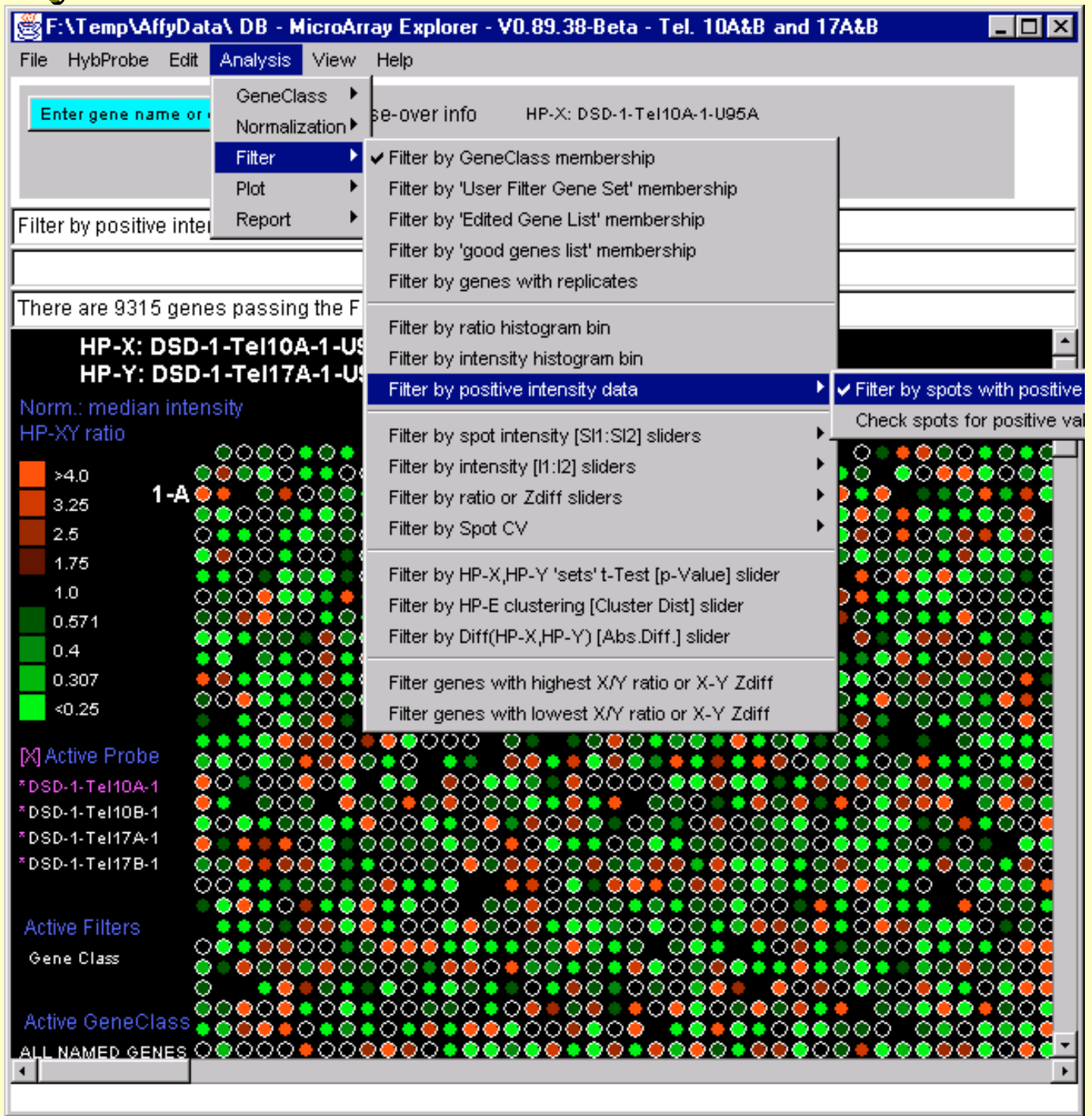
II.18b Incyte Data from Cvt2Mae Conversion



II.19a Converted Affymetrix files for Analysis



II.19b Affymetrix Data from Cvt2Mae Conversion



II.20 Using MAExplorer with mAdb Data

- The **NCI/CIT mAdb Web microarray database server** is an array data repository and analysis facility for microarrays created in conjunction with the NCI-ATC facility **(http://nciarray.nci.nih.gov/)**

- It can create a set of data files, downloaded as a Zip file from the mAdb, in a ready-to-use format compatible with MAExplorer

- **Section III** describes the procedure for **downloading MAExplorer.** You should periodically check the MAExplorer Web site to see if there is a major revision that you might want to download

- **Section IV** describes the procedure for **downloading a mAdb data set** and starting MAExplorer on that data.

- **Help desk for MAExplorer:** *mae@ncifcrf.gov*
  Peter Lemkin (301-846-5535)
  Greg Thornwall (301-846-5539)

Summary

- MAExplorer is a flexible microarray data-mining tool running on the user's computer
- Uses direct-manipulation, data filtering, built-in graphics, statistics, clustering, gene and sample set operations
- Manages multiple samples, replicates, sets, expression profile lists where the state may be saved on the disk for later use
- Cvt2Mae data conversion tools allows use with common chips
- Plug-ins will allow extension with new analytic methods by users
- MAExplorer identified genes in preferentially expressed during lactation, *Nucleic Acids Res.* (2000) **28**:4452
- Freely available for download with documentation on Web site